Proc. Natl. Acad. Sci. USA
Vol. 77, No. 12, pp. 7079–7083, December 1980
Biochemistry

Posttranscriptional modification of mRNA conformation: Mechanism that regulates erythromycin-induced resistance

(antibiotics/ribosomes/gene expression/*Staphylococcus aureus*/*Bacillus subtilis*)

SUEHARU HORINOUCHI AND BERNARD WEISBLUM

Pharmacology Department, University of Wisconsin Medical School, Madison, Wisconsin 53706

Communicated by Masayasu Nomura, August 25, 1980

ABSTRACT **The nucleotide sequence of a gene in plasmid pE194 responsible for erythromycin-induced resistance, including regulation of the resistance phenotype, is reported. A DNA fragment from plasmid pE194, obtained by digestion with *Taq* I restriction endonuclease, was cloned in *Bacillus subtilis* by using pC194 as the plasmid cloning vector. Erythromycin-resistant, inducible transformant clones containing the *Taq* I fragment A were obtained in which the expression of resistance was similar to that found in the original pE194 background; an interpretative model of the regulation of the erythromycin-resistance determinant is proposed based on the sequence of the *Taq* I A fragment. The cloned *Taq* I A fragment consists of 1442 base pairs and has open reading frames capable of coding for a peptide and a protein containing 19 and 243 amino acids, respectively, referred to as the "leader peptide" and "29,000 protein." Between the putative transcriptional start site and the ribosome binding site for 29,000-protein synthesis, the promoter region contains four complementary inverted repeat sequences named "1, 2, 3, and 4," respectively, in which 1 is complementary to 2, 2 is complementary to 3, and 3 is complementary to 4. Sequence 1 encodes the COOH-terminal half of the leader peptide, whereas the ribosome binding site for synthesis of 29,000 protein is sequestered in a loop formed by the association of 3 and 4. The 29,000-protein promoter region does not appear to contain any transcription stop signal. We propose a model for regulation of erythromycin resistance according to which ribosomes engaged in leader peptide synthesis are partially inhibited by optimal inducing (i.e., subinhibitory) concentrations of erythromycin that, in turn, cause an accumulation of these partially inhibited ("stalled") ribosomes in sequence 1. During induction, the translationally inactive states of association of the inverted repeats, postulated to be 1 plus 2 and 3 plus 4, respectively, are perturbed by a high level of stalled ribosome occupancy in sequence 1, and in the resultant redistribution, 2 associates with 3, freeing 4 and thereby freeing the ribosome binding site sequestered by the association of 3 and 4. Sequence alterations at the 5′ end of the 29,000-protein coding region associated with mutation to constitutive expression have been localized to the inverted complementary repeats, and determination of base changes in eight mutants are all capable of reducing the stability of the postulated stems in a manner consistent with predictions made by the model.**

The mechanism of erythromycin resistance in clinical bacterial isolates involves specific modification after synthesis—namely, N^6-dimethylation of adenine in 23S ribosomal RNA—which markedly reduces the affinity between erythromycin and the ribosome (1–3). Some properties of inducible strains and references to earlier work have been covered in a review (3). Cells exposed to 10–100 nM erythromycin for induction (optimal inducing concentration range) acquire resistance to concentrations exceeding 100 μM within 40–60 min, and in the course of induction the entire culture acquires the capability to express the resistance phenotype (2, 4). In the present study, data from preliminary cloning experiments suggested that regulation of resistance was exerted at a level other than transcription. Based on our experimental findings, we propose a mechanism of gene regulation in which an induced conformational change in the mRNA converts it from a form inactive in translation to one that is active as a result of complex formation between (*i*) erythromycin, (*ii*) sensitive ribosomes, and (*iii*) a regulatory sequence in mRNA.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Bacterial Strains and Plasmids. pE194 and pC194 are plasmids found originally in *Staphylococcus aureus* that specify resistance to erythromycin and chloramphenicol, respectively. Both plasmids can be introduced into *Bacillus subtilis* by transformation, during which they express their respective resistance phenotypes, and their use as cloning vehicles in *B. subtilis* has been described (5). pC194 DNA and pE194 DNA used for sequence determinations were prepared as described (6). Constitutive mutants of pE194, selected with tylosin, and their respective plasmid DNA preparations were obtained using *S. aureus* RN2442 (a strain that carries pE194) provided by R. P. Novick (7, 8). The recombinant experiments were performed using *B. subtilis* MI112 as recipient, as described by Tanaka (9, 10). pTL11, provided by T. Tanaka (10), served as source of the trimethoprim-resistance determinant.

Cloning the pE194 *Taq* I A Fragment Using pC194. pC194 contains five *Taq* I sites (recognition sequence: T-C-G-A) of which two belong to the *Cla* I subgroup (A-T-C-G-A-T), whereas pE194 contains three *Taq* I sites of which one belongs to the *Cla* I subgroup. The ligation mixture contained a limit pE194 *Taq* I digest and a partial pC194 *Cla* I digest in a 1:1 ratio. Transformants selected for erythromycin resistance were screened for chloramphenicol resistance. Covalently closed circular (form I) DNA was prepared from three selected transformants resistant to chloramphenicol and inducibly resistant to erythromycin. The purified form I DNA obtained from these strains was digested with *Taq* I, and the resultant digest was fractionated by polyacrylamide gel electrophoresis. This procedure yielded five fragments for pC194, six fragments for one of the recombinants (of which one band was indistinguishable in mobility from the pE194 *Taq* I fragment A), and seven fragments for the other two recombinants in which apparently both the *Taq* I A and B fragments were cloned. All enzymes used in these studies were purchased from New England BioLabs except for *Cla* I, which was purchased from Boehringer. The sequence of the *Taq* I A fragment was determined by the method of Maxam and Gilbert (11).

Sequence Alterations in Constitutive Mutants of pE194. A set of independent constitutive *S. aureus* mutants was selected by using enriched medium supplemented with 10 μg of tylosin

Abbreviation: S–D, Shine–Dalgarno sequence.

per ml. After locating the site of mutation near the 5′ end of the sequence that codes for the 29,000 protein by means of recombinant mapping techniques, sequence determinations were performed using DNA obtained from each of the mutants.

RESULTS AND DISCUSSION

Restriction Endonuclease Cleavage Sites in pE194. A map of pE194 listing the pertinent restriction endonuclease cleavage sites is shown in Fig. 1. Also indicated are genes coding for three proteins with molecular sizes of 15,000, 29,000, and 45,000 daltons deduced from the DNA sequence. They are referred to as the "15,000," "29,000," and "45,000" proteins, respectively. Experimental data pertaining to the 15,000 and 45,000 proteins will be presented elsewhere. The object of the present study was the region containing the sequence that codes for the 29,000 protein—the protein that mediates expression of the erythromycin-resistance phenotype.

Recombinant DNA Studies; Cloning the *Taq* I Fragments of pE194. Our most informative results were obtained from cloning a limit *Taq* I digest of pE194 by using a partial *Cla* I digest of pC194 as cloning-vehicle preparation. Erythromycin-resistant transformant clones were obtained containing the *Taq* I A fragment. One of these containing the *Taq* I A fragment alone showed inducible resistance when tested with disks containing either erythromycin and tylosin. These findings suggested that the *Taq* I B and C fragments of pE194 were not necessary for either resistance or its regulation, and that the *Taq* I A fragment alone could code for both the erythromycin-resistance determinant and its associated regulatory mechanism.

Insertional Inactivation of Erythromycin Resistance. The fragment chosen for insertion into the *Bcl* I site (recognition sequence: T-G-A-T-C-A) contained a trimethoprim-resistant dihydrofolate reductase determinant flanked by two *Bgl* II sites (A-G-A-T-C-T) from pTL11, which could be ligated, therefore, into the *Bcl* I site. Trimethoprim-resistant, erythromycin-sensitive transformant clones were obtained, and form I DNA prepared from one of these clones was analyzed by the heteroduplex method (12) and by polyacrylamide gel electrophoresis. The analyses showed that the trimethoprim-resistance determinant was inserted into the *Bcl* I site. We conclude that the 29,000 protein is necessary for expression of erythromycin resistance and that the *Bcl* I site interrupts the continuity of the coding sequence for this protein.

Localization of a Control Element for Erythromycin Resistance. To localize sequence alterations responsible for expression of the resistance phenotype, we prepared four recombinant plasmids using pE194 restriction fragments from inducible and constitutive pE194 DNA as outlined in Fig. 2. pE194 DNA was first cut with *Bcl* I and *Pst* I (each of which cuts pE194 once), followed by religation of the two complementary fragments from inducible pE194 and from a constitutive mutant. After transformation, erythromycin-resistant transformant clones were selected and scored with respect to whether resistance was inducible or constitutive. For the *Pst* I-*Bcl* I recombinant, this procedure allowed us to localize the determinant of inducible (or constitutive) expression in the smaller of the two fragments (i.e., between 0.68 and 0.07 map units, proceeding in a clockwise direction). A similar set of constructions, in which fragments prepared by digestion with *Msp* I were used, pointed to the region between 0.29 and 0.86 map units as the location of the regulatory determinant. These independent determinations narrow the site of mutation to the interval between 0.68 and 0.86 map units—an interval which covers the 5′ end of the sequence coding for the 29,000 protein. Finer localization of a possible promoter sequence in this region between 0.75 and 0.78 map units was suggested by the fact that

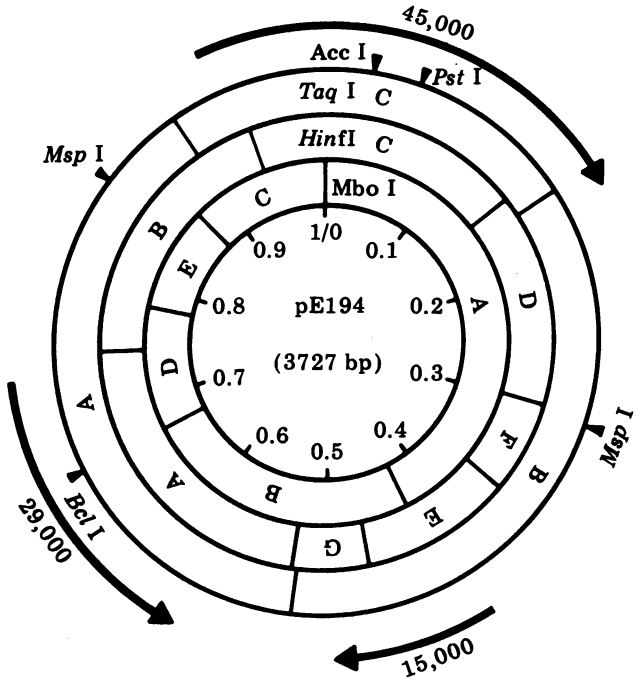

FIG. 1. Summary of restriction sites in pE194. The map is based on direct studies of digestion products with restriction endonucleases and confirmed by determination of the sequence. The three proteins designated are predicted by open coding sequences. bp, Base pairs.

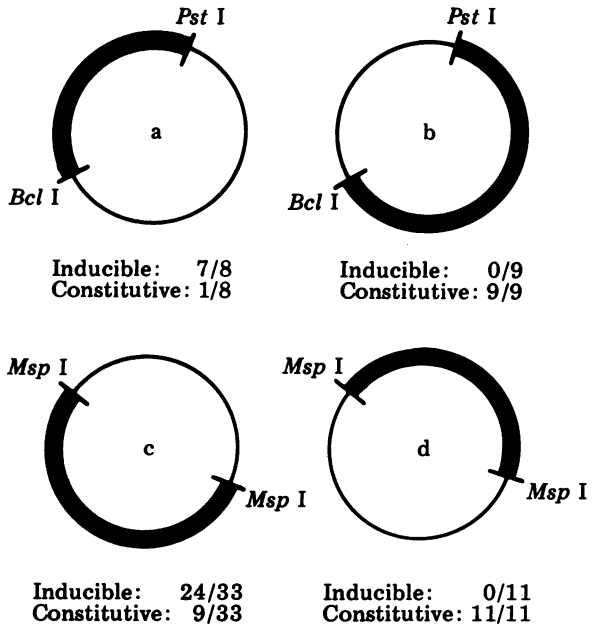

FIG. 2. Localization of the erythromycin control region. Heterologous plasmids were constructed from inducible and constitutive pE194 by using restriction endonucleases as indicated. After transformation into *B. subtilis*, erythromycin-resistant transformants were selected and scored with respect to whether their phenotype was inducible or constitutive. The enzymes used to obtain the fragments used in the constructions and the respective source phenotypes are indicated below. The thick segments were derived from inducible pE194 and the thin segments were derived from a constitutive mutant. (*a*) *Bcl* I–*Pst* I large (constitutive) + *Bcl* I–*Pst* I small (inducible). (*b*) *Bcl* I–*Pst* I large (inducible) + *Bcl* I–*Pst* I small (constitutive). (*c*) *Msp* I large (inducible) + *Msp* I small (constitutive). (*d*) *Msp* I large (constitutive) + *Msp* I small (inducible).

RNA polymerase formed a 0.6 M KCl-resistant complex in the absence of added triphosphates only with the *Hin*f I B and *Mbo* I D fragments by using the method of Taylor and Burgess (13) (data not shown). The finer localization represents the small overlap between the two fragments.

Nucleotide Sequence of the *Taq* I A Fragment. In the results presented below, "residue" refers to an absolute enumeration based on the sequence of the entire plasmid (3727 base pairs). In view of the noteworthy biological properties of the *Taq* I A fragment, we report its nucleotide sequence (Fig. 3). The DNA sequence contains two open reading frames capable of coding for two peptides containing 19 and 243 amino acids, respectively; the latter corresponds to a protein with a calculated molecular size of 29,000 daltons. Immediately preceding the 29,000-protein coding sequence we note a region that resembles promoters described in other systems; an interpretive summary of the properties of this region is shown in Figs. 3 and 4. Preceding the initiator codon at residue 2738 we note a series of inverted complementary repeat sequences, and at residue 2857 we note the beginning of a sequence capable of coding for a polypeptide that contains 19 amino acids. This short coding sequence is preceded by a Shine–Dalgarno (14) sequence (S–D-1), a Pribnow sequence (PB-1), and a potential RNA polymerase recognition site (-35-1). An extensive listing of promoter sequences associated with the name-designations used, including references to earlier studies, has been reviewed by Rosenberg and Court (15). The two adenines at residues 2879 and 2878 would be situated at a suitable site to serve as potential A-starts for the initiation of transcription. Between the promoter region and the 29,000-protein initiator methionine at residue 2738, we note two pairs of complementary inverted repeats, the first of which (labeled 1 and 2) overlaps the COOH-terminal portion of the sequence that codes for the 19-amino acid peptide referred to above. The second pair (labeled 3 and 4) overlaps an additional set of possible promoter-associated sequences as indicated, including a possible RNA polymerase binding site (-35-2), a PB-2 and a S–D-2. The S–D-2 localized within the loop formed by the second pair of inverted repeat sequences plays a key role in the model proposed below.

At the 3′ end of the putative mRNA transcript, sites for 29,000-protein translation termination and transcription termination are also indicated (Fig. 3). The transcription terminator contains a complementary inverted repeat followed by oligo-(T) similar to that found in other systems (15–17).

A Model for Regulation of Resistance. The two pairs of inverted complementary repeat sequences 1 plus 2 and 3 plus 4 described in Fig. 3 are shown with additional interpretative details in Fig. 4. Because mutations in this region alter the rate

FIG. 3. Nucleotide sequence of the *Taq* I A fragment. The origin for numbering the nucleotide residues is located at the *Mbo* I A–C boundary as shown in Fig. 1. The high residue numbers reflect this fact, and their descending order in proceeding from promoter to gene reflects the fact that the 29,000 protein is read in both the reverse orientation and opposite strand relative to the 15,000 and 45,000 proteins. The *Bcl* I site mentioned in the text begins at residue 2520, and the *Hin*fI site used for mutant sequence determinations begins in stem 2 at residue 2786.

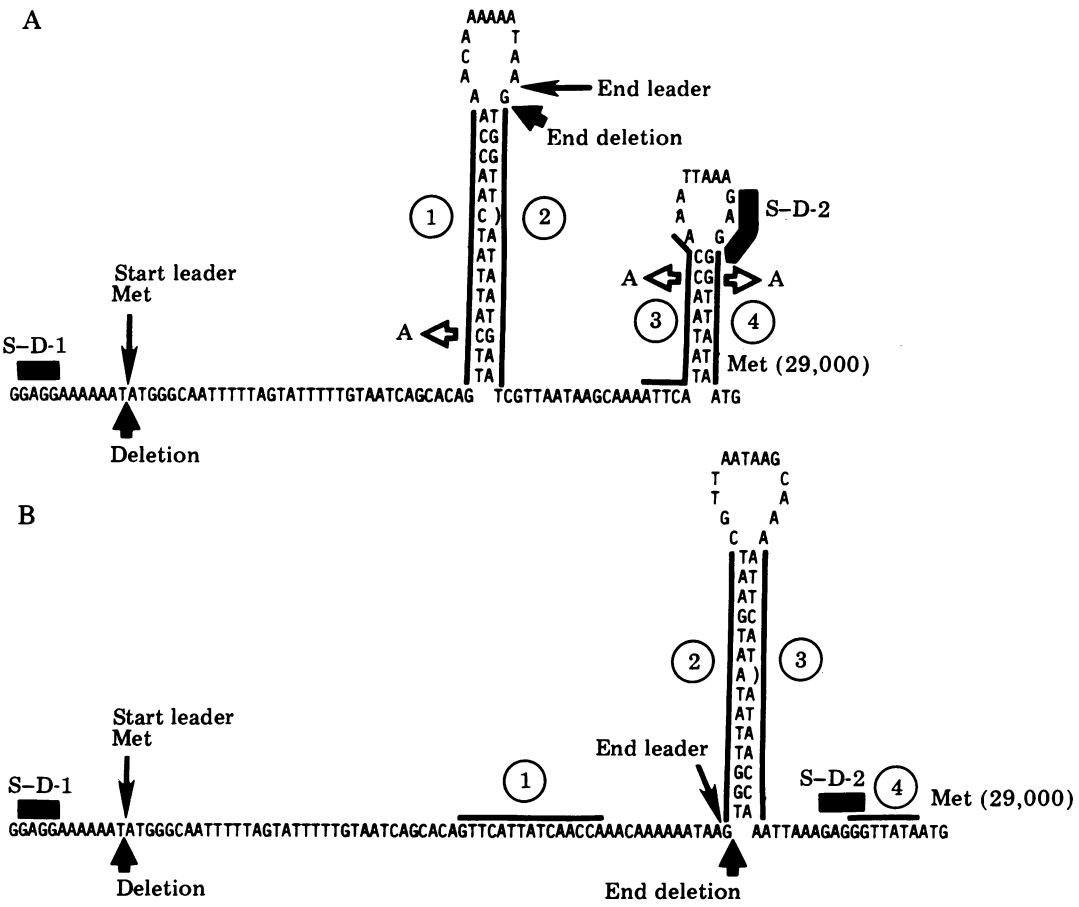

FIG. 4. Schematic of alternative stem combinations in the 29,000-protein control region. (*A*) Inactive conformation showing sequestration of S–D-2 in the loop formed by association of 3 and 4. Sites of point mutations to constitutivity are indicated with open arrows. For details see text. (*B*) Active conformation in which S–D-2 has been freed, making it available for ribosome attachment and synthesis of the 29,000 protein. The stems shown are postulated to correspond to active and inactive conformations of the control region at the level of mRNA. Ts have been used instead of Us to facilitate comparison with the DNA sequence shown in Fig. 3 and discussed in this paper. The labels used correspond to those used in Fig. 3.

of expression of 29,000-protein function, we must explain the regulation of expression in terms of the structure of this region and propose the following model to account for our observations. The association of sequences 1 plus 2 and 3 plus 4 in mRNA forms one of several possible sets of base-paired stem combination; we postulate (*i*) that this conformation corresponds to the low rate of expression since S–D-2 is trapped by 3 plus 4 stem formation, and (*ii*) that disruption of the integrity of 3 plus 4 to free up S–D-2 constitutes the key event in induction. We postulate further (*iii*) that during induction, ribosomes first bind to S–D-1 to initiate synthesis of the leader peptide. In the presence of subinhibitory inducing concentrations of erythromycin, the ribosomes become "stalled" in sequence 1; as a consequence, 2 is diverted to form an energetically favored association with 3, which frees 4 and therefore S-D-2, which can then bind ribosomes permitting them to initiate synthesis of the 29,000 protein. The control region that we have sequenced shows some similarity to attenuator regions associated with operons responsible for the biosynthesis of amino acids. Some of the properties of these attenuator regions in various systems have been discussed in reviews of the subject (16, 17). In contrast, the erythromycin control region contains no obvious transcription termination signal; therefore, we infer that the function of this region differs from that of the attenuators reviewed above.

Mutations in the Control Region: Supporting Evidence for the Model. Several mutants that we have analyzed support this model. Eight independent constitutive *S. aureus* mutants selected using tylosin were randomly chosen for sequence analysis. The distribution of these mutants is shown in Fig. 4; three separate independent occurrences of C to A in 3, and G to A in 4 point to the importance of the role of 3 plus 4 pairing stability in maintaining the repressed state. The involvement of a single C·G base pair in the 3 plus 4 stem in six of eight independent mutants suggests that these two distinct mutable residues, separated by 13 bases, have biological significance as a base pair *in vivo*; loss of either the C or the G at this critical stem site would also result in loss of the integrity of the C·G base pair immediately adjacent to the loop, owing to a cooperative effect and favoring dissociation of 3 plus 4. We found one constitutive mutant in which the entire sequence 1 and the 5′ end of the sequence that codes for the leader peptide (61 bases) were deleted as indicated in Fig. 4. In this deletion mutant, any negative regulatory effects exerted by 1 plus 2 pair formation would be bypassed, according to the model proposed. Finally, one occurrence of a point mutation (C to A) was found in 1, which would serve to weaken its association with 2. The pivotal role postulated for 2 requires that its integrity be maintained in mutation to constitutivity, whereas the opposite is true for 1, 3, and 4; in agreement with this expectation, none of the mutants found so far involves sequence 2.

Refinements to the Model for Induction. Several questions concerning the model may be asked. (*i*) If induction requires ribosome stall, should not the same ribosome stall also prevent

synthesis of the 29,000-protein as well? (*ii*) If inhibition of ribosome function provides the inducing stimulus, why cannot any inhibitor of ribosome function induce resistance to erythromycin?

In response, it is assumed that repression is not 100% effective and that the uninduced cell normally contains a small number of 29,000-protein molecules and therefore resistant ribosomes. We postulate that during induction, the majority sensitive ribosomes are responsible for ribosome stall, whereas the minority resistant ribosomes synthesize the 29,000 protein once they have bound to S-D-2. Such a reaction, slow at first, would become autocatalytic once a critical concentration of resistant ribosomes is attained. This would be consistent with the observations of Weaver and Pattee (4) that inducible cells grown in the absence of antibiotic, innoculated into challenging medium containing 100 μg of erythromycin per ml (about 100 times the minimum inhibitory concentration) eventually started growing after a prolonged lag of about 10 hr. (Cells preinduced by exposure for 15–30 min to 0.1 μg of erythromycin per ml, about 0.10 μM, grew with a lag of about 2 hr and reached maximum growth by 8 hr.) This observation is consistent with the notion that either (*i*) stalled ribosomes can eventually traverse the messenger to make a functional protein, or (*ii*) repression is not 100% effective and the uninduced cell has a small number of resistant ribosomes. Finally, the expression of resistance would turn itself off once the concentration of resistant ribosomes exceeded a critical threshold, owing to the requirement for sensitive ribosomes to disrupt 1 + 2. In the context of negative feedback features of this model, it also may be possible for 1 to associate with 4 in the absence of an inductive stimulus with resultant masking of S-D-2 as well as the codons for Met and Asn in the 29,000 protein. Moreover, we must consider the possibility that the 29,000 protein may serve as a repressor of its own synthesis at the translational level by binding to and stabilizing 3 + 4. The sequence A-A-A-G present in the loop formed by 3 + 4 is identical to the methylated sequence in 23S ribosomal RNA from *S. aureus* (3), a site presumed to be recognized by the 29,000 protein. The occurrence of similar sequences at both the site of methylation in the rRNA and the site for control of expression in the mRNA could serve as an additional feature of this regulatory mechanism.

Cundliffe (18), Pestka (19), Tanaka and Teraoka (20), and Tai *et al.* (21) have made detailed studies of inhibition produced by erythromycin. They noted that erythromycin can inhibit at an early stage after initiation, before many amino acid residues have been incorporated. Vazquez (22) in a review has concluded that the inhibitory action of erythromycin "is preferential for polysomes synthesizing short (di- to penta-) peptides, and the efficiency of inhibition depends on the amino acid composition of the peptide." In the context of the model for inducible resistance, this is precisely the type of inhibitory action required for an antibiotic to have significant inducing activity. How the pattern of codon utilization in the leader peptide optimizes its function in induction, or whether the leader peptide performs some specific function is not yet clear; however, the small size of peptides synthesized in an erythromycin-inhibited system compares favorably with the peptide length needed (12–15 residues) to occupy sequence 1 and to prevent formation of a double-stranded complex with sequence 2. The significant attributes of the macrolide, lincosamide, and streptogramin type B antibiotics relevant to their effectiveness as inducers may be that (*i*) it is only the macrolide, lincosamide, and streptogramin type B antibiotics to which the minority population of methylated ribosomes postulated to be normally present is resistant, and (*ii*) the macrolide, lincosamide, and streptogramin type B antibiotics may inhibit ribosome function in a way consonant with ribosome stall in the leader peptide region.

We are grateful to L. Post, P. O. Olins, and M. Nomura, and to N. Panayotatos and R. D. Wells for instruction in the art and science of DNA sequence determinations. M. Jacobson, T. Baisch, J. Bacon, and P. Jaszewski provided expert technical assistance in sequence determination. We thank W. Taylor and R. B. Burgess for assistance in the RNA polymerase binding studies; M. Fiandt and W. Szybalski for helping us in mapping the location of the trimethoprim-resistant dihydrofolate reductase gene that was kindly provided by T. Tanaka. E. Cundliffe, W. Reznikoff, and I. Tessman provided enthusiastic criticism and many helpful suggestions. This work was supported by a research grant from The National Science Foundation (PCM-7719390), as well as by grants from the Research Foundations of The Upjohn Company and Eli Lilly and Company.

1. Lai, C.-J. & Weisblum, B. (1971) *Proc. Natl. Acad. Sci. USA* **68,** 856–860.
2. Weisblum, B., Siddhikol, C., Lai, C.-J. & Demohn, V. (1971) *J. Bacteriol.* **106,** 835–847.
3. Weisblum, B. (1975) in *Microbiology-1974,* ed. Schlesinger, D. (ASM Publications, Washington, DC), pp. 199–206.
4. Weaver, J. R. & Pattee, P. A. (1964) *J. Bacteriol.* **88,** 574–580.
5. Gryczan, T., Shivakumar, A. G. & Dubnau, D. (1980) *J. Bacteriol.* **141,** 246–253.
6. Weisblum, B., Graham, M. Y., Gryczan, T. & Dubnau, D. (1979) *J. Bacteriol.* **137,** 635–643.
7. Iordanescu, S. (1976) *Arch. Roum. Pathol. Exp. Microbiol.* **35,** 111–118.
8. Iordanescu, S. (1977) *J. Bacteriol.* **129,** 71–75.
9. Tanaka, T. (1979) *J. Bacteriol.* **139,** 775–782.
10. Tanaka, T. & Kawano, N. (1980) *Gene* **10,** 131–136.
11. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
12. Westmoreland, B. C., Szybalski, W. & Riis, H. (1969) *Science* **163,** 1343–1348.
13. Taylor, W. E. & Burgess, R. R. (1979) *Gene* **6,** 331–365.
14. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 1342–1346.
15. Rosenberg, M. & Court, D. (1980) *Annu. Rev. Genet.* **13,** 319–353.
16. Johnston, H. M., Barnes, W. M., Chumley, F. G., Bossi, L. & Roth, J. R. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 508–512.
17. Keller, E. B. & Calvo, J. M. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 6186–6190.
18. Cundliffe, E. (1969) *Biochemistry* **8,** 2063–2066.
19. Pestka, S. (1974) *Antimicrob. Agents Chemother.* **5,** 255–267.
20. Tanaka, K. & Teraoka, H. (1966) *Biochim. Biophys. Acta* **11,** 204–206.
21. Tai, P.-C., Wallace, B. J. & Davis, B. D. (1974) *Biochemistry* **13,** 4653–4659.
22. Vazquez, D. (1975) in *Antibiotics III. Mechanism of Action of Antimicrobial and Antitumor Agents,* eds. Corcoran, J. W. & Hahn, F. E. (Springer, New York), pp. 459–479.